



# CERES Data Management Group Status Report

CERES Science Team Meeting

April 26<sup>th</sup>, 2016

Jonathan Gleason



# Data Management Team

**Systems/  
Optimization:**  
Nelson Hillyer  
Josh Wilkins  
Mike Linsinbigler

**Production /  
Optimization:**  
Lisa Coleman  
Carla Grune

**Instrument:**  
Denise Cooper  
Thomas Grepiotis  
Dianne Snyder

**Convolution:**  
Igor Antropov

**Inversion:**  
Victor Sothcott

**Product Web  
Support:**  
Churngwei Chu  
Ed Kizer  
Cristian Mitrescu

**Configuration  
Management:**  
Tammy Ayers  
Joanne Saunders

**ERBE-like:**  
Dale Walikainen  
Jeremie Lande

**SARB:**  
Tom Caldwell

**Task Manager:**  
Walt Miller

**Clouds:**  
Sunny Sun-Mack  
Ricky Brown  
Yan Chen  
Liz Heckert  
Rita Smith  
Sharon Gibson

**TISA Gridding:**  
Raja Raju  
Forrest Wrenn

**TISA Averaging:**  
Cathy Nguyen  
Dennis Keyes



# Overview

- Recent DMT Activity
- Current Data Availability
- CATALYST Status
- Discontinuation of Beta2-Ed4 SSFs
- Delivery tracking using JIRA
- Systems Status



# DMT Recent Activity



# Recent Activity

## 43 Software & data deliveries since September 1<sup>st</sup>, 2015

- **CATALYST (12)**
  - Server & Ops Console version 2.0
  - Ed4 SSF, SYN1deg & SSF1deg PGE Handlers
  - Logging DB patches & updates
- **PR Tool (6)**
  - CATALYST update support
  - New FluxByCloud PGEs
- **Instrument (3)**
  - 2 PGE updates for NPP Special Ops
- **CERESLib (1)**
  - New SOFA ground validation sites
- **Clouds (1)**
  - PCF update to handle “no aerosol” case
- **Inversion (5)**
  - Read SSFi vs Beta2-Ed4 SSF as input
  - Added SOFA ground validation site
- **TISA Gridding (3)**
  - NPP Edition 1 updates
  - 3 New Ed4 CldTypHist PGEs
- **TISA Averaging (7)**
  - Ed3 SSF1deg redelivered to Use GMT and Ed4 format
  - Ed4 SSF1deg correction & x86 only
- **SARB (4)**
  - Ed4 SYN1 code updates
  - Delta MATCH file delivery



# Data Availability



# Edition 3 Product Availability

Product	Platform	Processed through	Publically Available
BDS	Terra & Aqua	December 31, 2015	Yes
SSF	Terra & Aqua	December 31, 2015 (FM1 & FM3 only)	Yes
SFC	Terra & Aqua	November 2015	Yes
SSF1deg-Month SSF1deg-Day	Terra & Aqua	November 2015	Yes
SYN1deg	Terra + Aqua	November 2015	Yes
ISCCP-D2like-Day/Nit	Terra & Aqua	December 2015	Yes
ISCCP-D2like-GEO	GEO	December 2015	Yes
ISCCP-D2like-Mrg	Terra + Aqua + GEO	December 2015	Yes
SSF Edition 1-CV	Terra & Aqua	February 29, 2016	No



# Edition 4 Product Availability

Product	Platform	Processed Through	Expected Current	Publically Available
BDS	Terra & Aqua	December 31, 2015	November 2015	Yes
SSF Beta2-Ed4	Terra & Aqua	December 31, 2014	N/A	No
SSF	Terra & Aqua	December 31, 2014	May 2016	Yes
SSF1deg-Hour	Terra & Aqua	November 30, 2014	May 2016	Yes
SSF1deg-Day/Month	Terra & Aqua	November 30, 2014	May 2016	Yes
SYN1deg	Terra + Aqua	July 2005 – July 2012	July 2016	No
CldTypHist	Terra + Aqua	Start June 2016	August 2016	No





# Edition1 Product Availability

Product	Platform	Processed Through	Expected Current	Publically Available
BDS	S-NPP	December 31, 2015	March 2016	Yes
SSF		December 31, 2015	March 2016	Yes
SSF1deg-Hour		March 2014	May 2016	No
SSF1deg-Day/Month		March 2014	May 2016	No



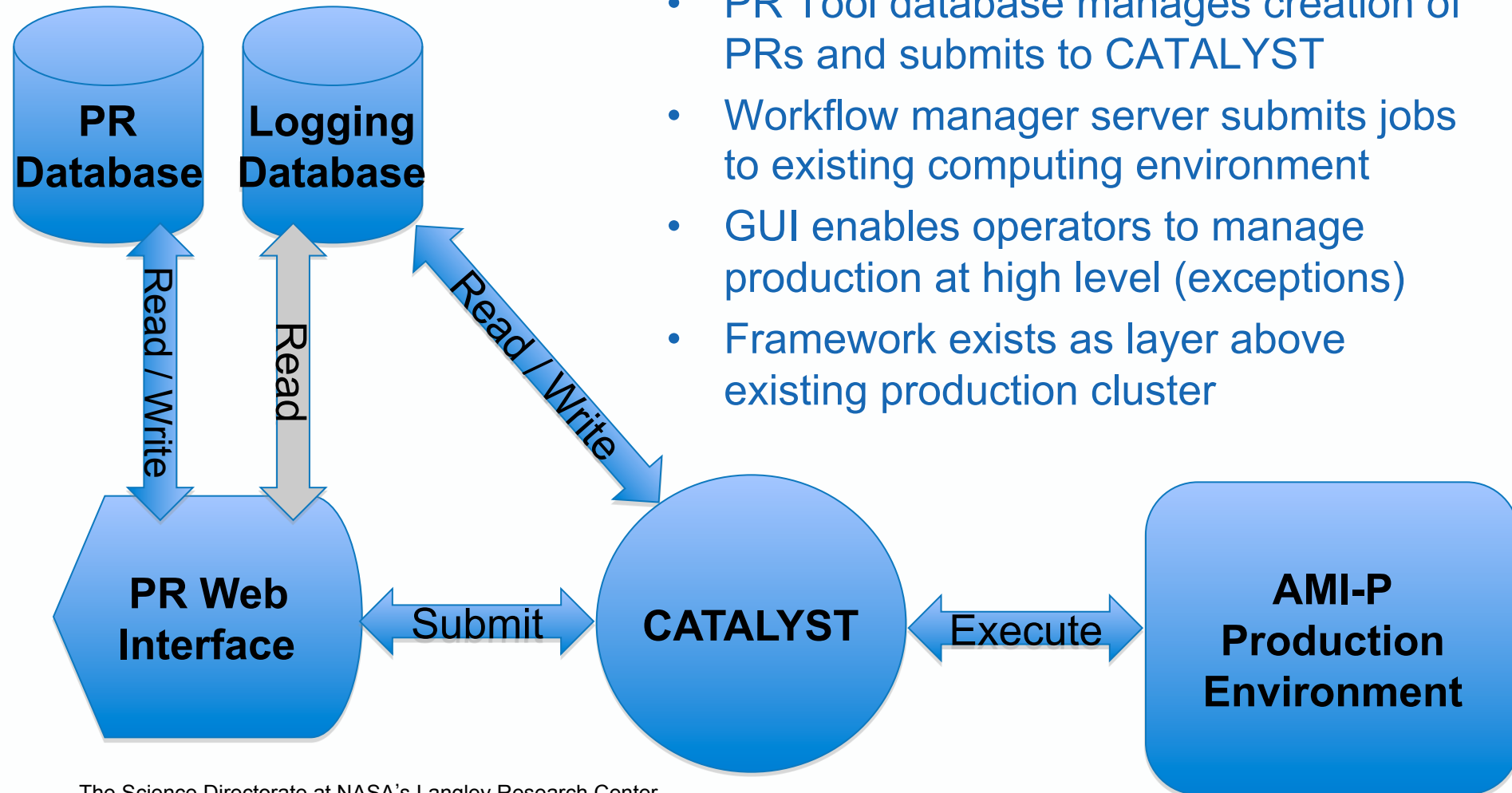
# CATALYST Status



# Automation Framework

## CERES AuTomAted job Loading sYSTem (CATALYST)

- PR Tool database manages creation of PRs and submits to CATALYST
- Workflow manager server submits jobs to existing computing environment
- GUI enables operators to manage production at high level (exceptions)
- Framework exists as layer above existing production cluster





# CATALYST Status

## CATALYST Release 1.0 supported Clouds and Inversion PGEs only

- Processed Beta2-Ed4 SSFs through December 2014 for Terra & Aqua
- Successfully ran ~505,448 jobs (328 data months)

## CATALYST Release 2.0

- Implements user requested features for increased usability
- Included significant server improvements for maintainability and sustainability
- Version 2.0 promoted live January 5<sup>th</sup>, 2016
- Migration of remaining Edition 4 PGEs into CATALYST in progress

# CATALYST PGEs By Stream





# PGE Migration Plan

## Edition 4 SSF Stream: Completed April 2016

- 4.5-6.0P6
- 4.5-6.1P6
- Handler testing enabled discovery of PGE issues early in process

## Edition 4 SYN1deg Stream: Delivered, Expected Complete May 2016

- 7.3.1P2, 7.3.1P3, 7.3.1P4
- 7.2.1P2
- 8.1P2

## Edition 4 SSF Stream: Delivered, Expected Complete June 2016

- 9.2P3, 9.3P3, 9.4P3
- 10.0P5



# PGE Migration Plan

MOA / PMOA / SAH: Deliver June 6<sup>th</sup>

- 12.1P2
- 9.1P2
- 5.0P3

Ed4 / NPP Ed1 Instrument & ERBElike: Deliver July 8<sup>th</sup>

- 1.2P1, 1.4P3
- 2.2P1, 2.3P1, 2.3P2, 2.4P1
- 3,1P2, 3.2P1

CATALST v2.1 software updates: Deliver August 12<sup>th</sup>

- Forward processing support
- Invert Logging DB to “known Missing only” format



# PGE Migration Plan

Edition 1 NPP SSF Stream: Deliver September 30<sup>th</sup>

- 4.1-4.1P7
- 4.1-4.2P6
- 4.1-4.2P7
- 4.1-4.3P4
- 4.5-6.1P7

Baseline1-QC & Edition1-CV Forward Streams: Deliver October 28<sup>th</sup>

- 1.0P3
- 1.1P8
- 1.4P1
- 1.4P2

Snow & Ice Forward Processing PGEs: Deliver December 9<sup>th</sup>

- 2.1P1
- 4.1-4.0P2
- 4.1-4.0P3





# Discontinue Beta2-Ed4 SSFs



# Edition 4 SSFs

To Synchronize Edition 4 Clouds and ADMs, required interim Beta2-Ed4 clouds product for ADM development

- Initially processed Ed4 Clouds with Beta-Ed4 Inversion
- Edition 4 Inversion code delivered to read Beta2-Ed4 SSF and Edition 4 IES as input

Starting at January 2015 Edition 4 Inversion code will directly read SSFi from Clouds

- Beta2-Ed4 Inversion code discontinues
- Result: Only one Edition 4 SSF data set will exist after 2014

This change timed with transition of Edition 4 Inversion PGEs into CATALYST



# Delivery Tracking Using JIRA



# CERES Deliveries In JIRA

Using JIRA to track CERES deliveries and communicate between ASDC and DMT

- SCCR high level workflow – 1 ticket per SCCR
- Each SCCR ticket composed of 1 or more “sub-tickets” of one of 5 different types
  - PGE Delivery - 19 to 25 possible workflow steps
  - Delta Script Delivery
  - Delta Data Delivery
  - Library Delivery (CERESLib / Perl\_LIB)
  - Automation Delivery (CATALYST or PR Tool)
- JIRA provides dashboard tool for customized reporting



# CERES Deliveries In JIRA

Have successfully used this system for last 35 deliveries

- JIRA tickets replace email for communication
- Workflow handles transition of ownership between DMT and ASDC (Ex: when ValR approval needed)
- System open for all with JIRA account to see status of any ticket
- Option exists to create “public” dashboard pages for WG leads to view status for deliveries of interest

Plan to transition ticket ownership to DMT Subsystem Leads in May

- SS Developer will track tickets
- ValR evaluation directly linked to subsystem
- DMT Lead not required middle person

<https://asdcjira.larc.nasa.gov:8443/secure/Dashboard.jspa>



# Production Status Reporting

## Production status available in Confluence

- Monthly priorities from DMT
- Weekly status updates from Ops

## Click below link to view (Open within LaRC firewall)

- “DMT Production Information”: Links to page of all monthly processing priorities provided
- “OPS Processing Status”: Links to list of months, each month links to page of weekly updates for that month

<https://asdcjira.larc.nasa.gov:9443/pages/viewpage.action?pagelId=18352593>

Started a Spring Science Team Meeting processing priorities dedicated page in January 2016 (time to prepare for Fall Meeting)



# Systems



# Systems Status

Production environment composed of P6, x86 and supplemented with IBM P7 blades

- Currently NPP Clouds runs on P7 blades to distribute load
- Ed3 & Ed4 SYNI & all Inversion run on x86
- Edition 4 Clouds runs on P6

Broad distribution of blade types helps for now by avoiding contention among high-resource consuming parallel processing streams

Tentative long-term plan to migrate exclusively x86 model





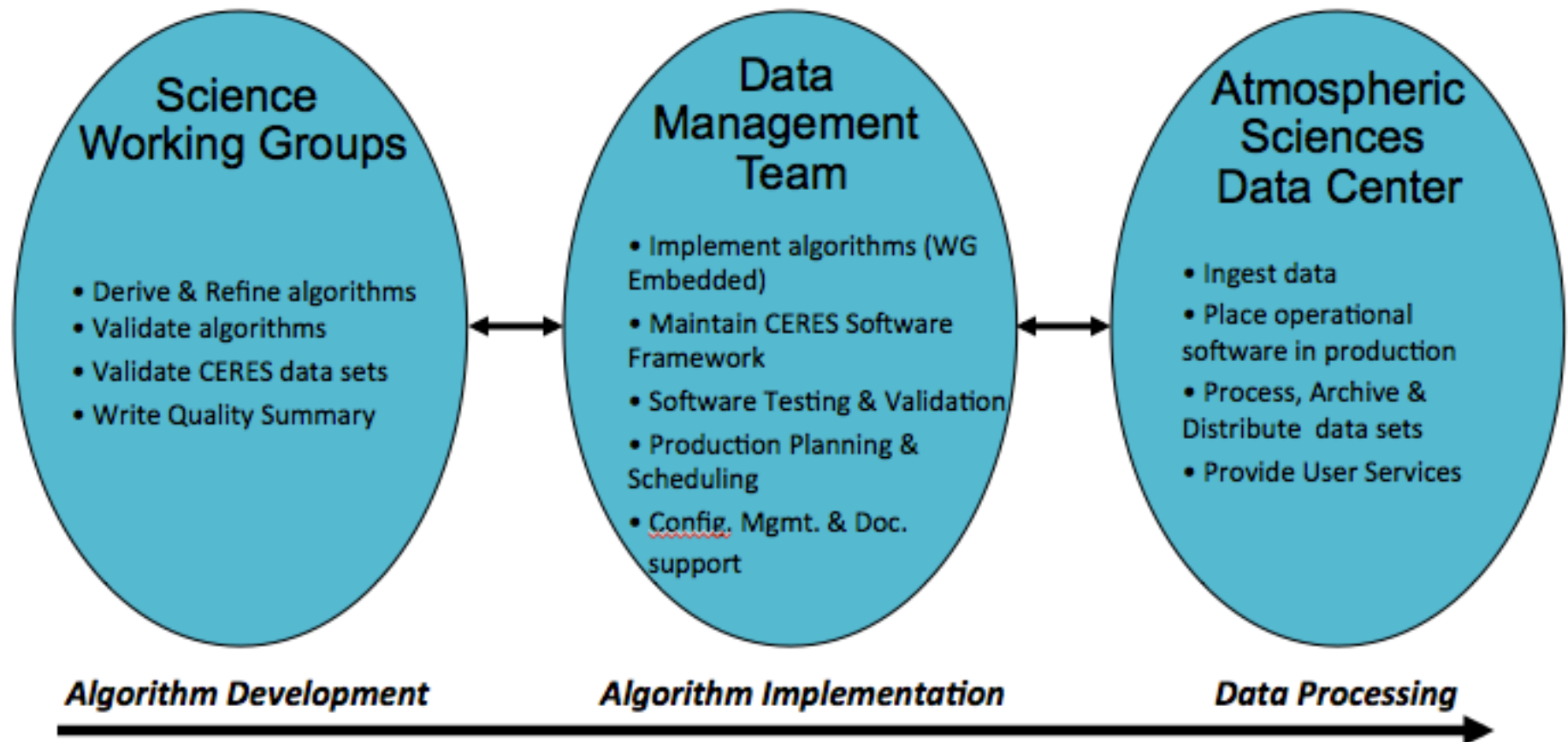
# Backup Slides

# CERES Production Software Overview



Subsystem Number	Subsystem Name	Number of PGEs	LOC (to nearest 1K)	Publicly Available Data Products	Product Frequency
1	Instrument/Pre-Processor	2	9K		
1	Instrument	6 + lib	162K	BDS	1/day
2	ERBE-like/ Inversion	5 + lib	31K	ES-8	1/day
3	ERBE-like/ TSA	2	14K	ES-9, ES-4	1/month
4.1 – 4.4	Clouds/VIIRS Subset Code	1	21K		12/hour
4.1 – 4.4	Clouds	11	358K		
4.5 – 4.6	Inversion	11	110K	SSF	1/hour
5	SARB	1	36K		1/hour
6 & 9	TISA-Gridding	7	34K	SSF1deg-Hour, ISCCP-D2like-Day/Nit	60/month, 36/month, 1/month
11	GGEO	1	6K	ISCCP-D2like-GEO	1/month
7.2	Synoptic SARB	1	47K		
7.1 & 8 10	TISA-Averaging	4	102K	SSF1deg-Day, SSF1deg-Month, SYN1deg-(3Hour, M3hour, Mhour, Month)	1/day, 1/month, 1/month, 5/month
12	MOA	2	14K		
	CERESlib/Perl_Lib		131K		
	<b>Total</b>	<b>54</b>	<b>1,075K</b>		

# Data Management Function



Currently ~1M production LOC contained in 54 active PGEs



# NPP Data Flow Changes

Starting in September 2015, CERES FM5 data for NPP will no longer be obtained through NOAA IDPS

- EDOS to connect to JPSS ground station stream & package to EOS Level-0 format files
- All beta testing for CERES complete – Using small pre-processor PGE to merge 24-hour CERES file with 2-hour Attitude and Ephemeris files into 24-hour CERES files
- IDPS will no longer process RDRs

This approach anticipated to be used for FM6 on J1 and for RBI

**Mention VIIRS new formats**



# CATALYST 2.0 Improvements

## 38 Change Requests submitted based on version 1.0

- 35 directly implemented or addressed in release 2.0
- 1 identified as too difficult to implement
- 2 delayed, likely to be superseded by release 2.0 capabilities

## Improvements by category

- PR Validation
- Problem Isolation
- PGE Management
- Operator's Console Improvements
- Stability & Monitoring
- Logging
- Security



# CATALYST 2.0 Improvements

## PR Validation

- Extensive PR validation logic added – verify invalid or conflicting submitted PRs

## Problem Isolation

Separate CATALYST to 6 sub-processes + Master process

- Limits client application downtime (i.e. PR Tool & Ops Console)
- Enables better logging and status reporting
- Improves Stability – Master processes monitors component health
- Contains known exception cases (i.e. server crashes, offline database, offline SGE)



# CATALYST Build 2.0

**Kernel** – PR validation & coordination (PGE Handlers live here)

**Cluster** – Interfaces with SGE

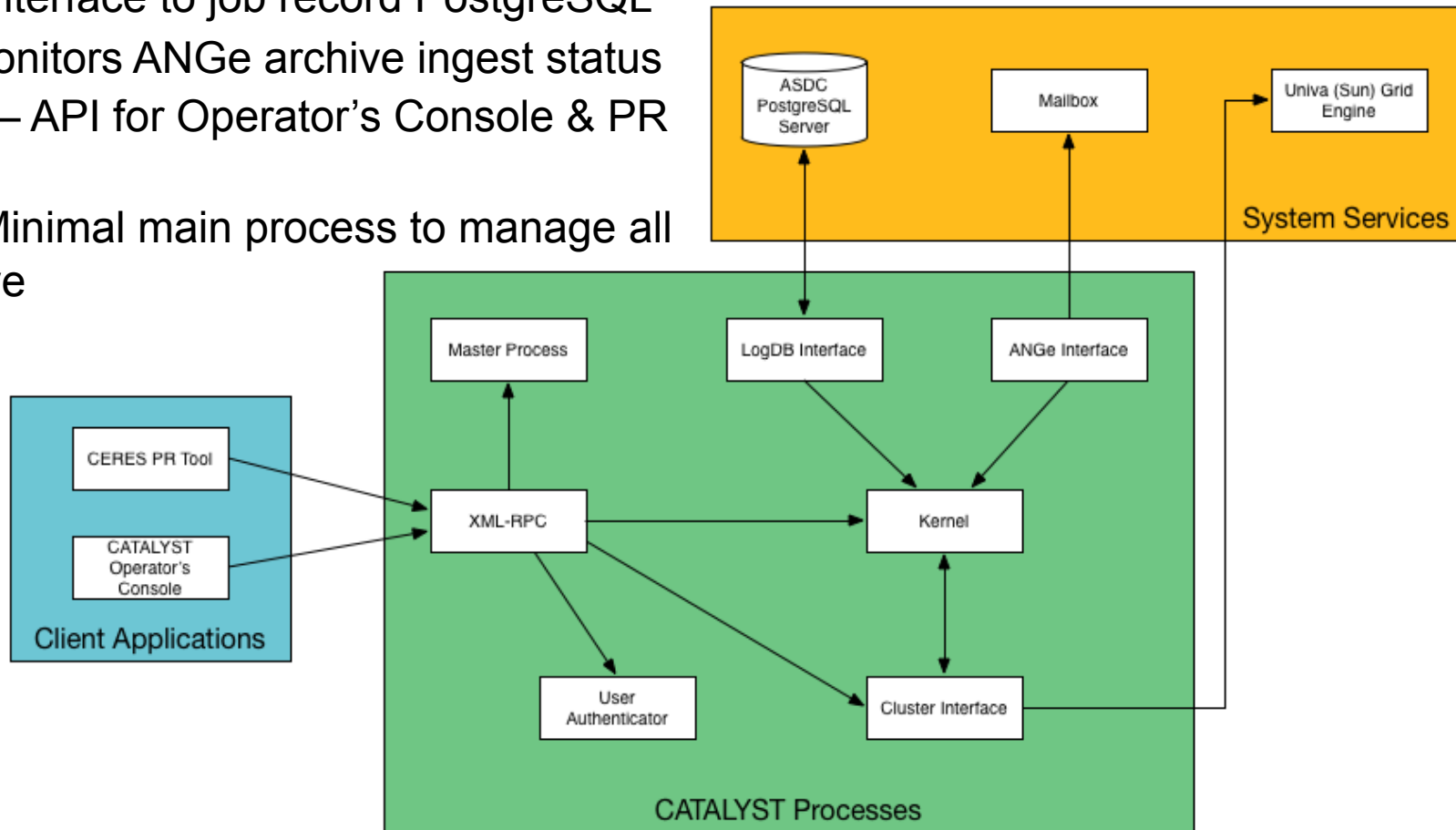
**User** – Authenticates users against LDAP

**LogDB** – Interface to job record PostgreSQL

**ANGe** – Monitors ANGe archive ingest status

**XML-RPC** – API for Operator's Console & PR Tool

**Master** – Minimal main process to manage all of the above

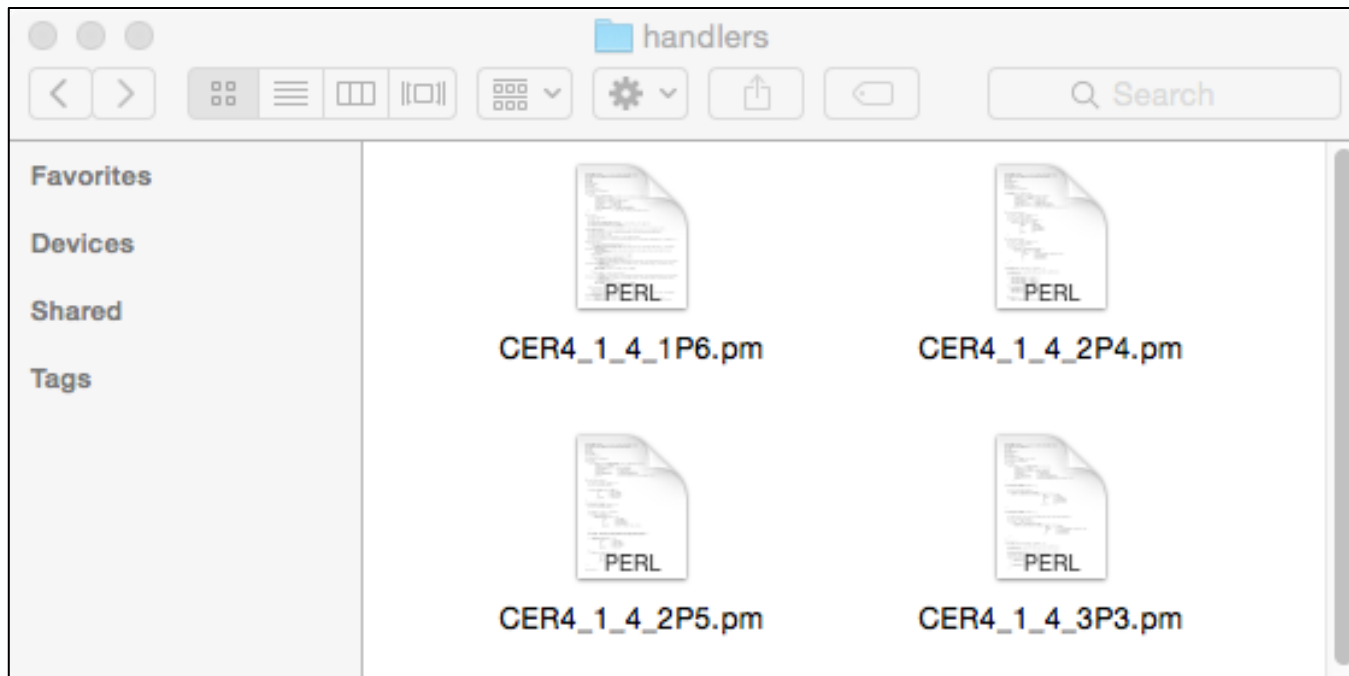




# CATALYST 2.0 Improvements

## PGE Management – file level

- Each PGE implemented as individual Perl module
- Modules installed in “CATALYST\_HOME/handlers” directory
- PGE modules loaded on start of CATALYST kernel process
- Add/Remove PGE achieved by copy/delete .pm and restart kernel process







# CATALYST 2.0 Improvements

## PGE Management – Within Operator's Console

- Whole Subsystem or individual PGEs can be enabled or disabled within UI.
- Epilogue settings can be adjusted by subsystem or PGE
- PGE and Epilogue settings modifiable only by privileged CATALYST user (viewable by all)
- Epilogues run in FIFO order – if predecessor disabled follow-on will wait (prevents out of order file removal)

PGE Settings

Subsystem: Clouds

Subsystem Mode: ☒ Enabled ☐ Disabled

Subsystem Epilog Mode: ☐ Enabled ☒ Disabled

PGE: CER4.1-4.2P4

PGE Mode: ☒ Enabled ☐ Disabled

PGE Epilog Mode: ☐ Enabled ☒ Disabled

PGE Details:

available -> on  
description -> Edition4 Daily QC Processor  
granularity -> CATALYST::DataDate::Daily

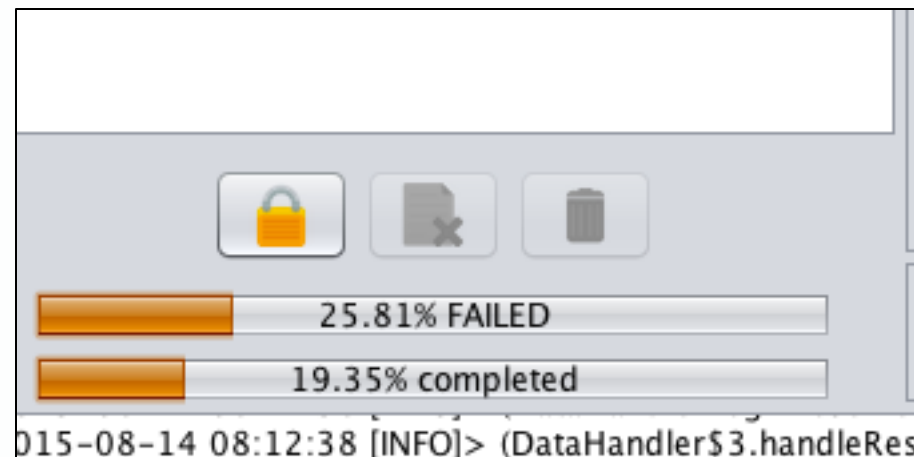
Apply Changes Close



# CATALYST 2.0 Improvements

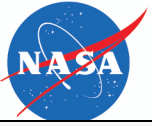
## Operator's Console Improvements

- Added text search capability – search by date, job status, epilogue status or job ID
- Job status reporting
  - Added progress bars for jobs completed/failed
  - Color coded job listings (complete, in progress, failed, waiting)
- Progress bars and activity logs for user actions
- Read-Only Mode replaces need for v1.0 recovery utility – when server started, all job status' viewable before administrator initializes jobs
- Inline documentation with hovers text



Percentage failed/complete

# CATALYST 2.0 Improvements



CATALYST Operator's Console (Connected - Server Uninitialized)

File CATALYST Server View Tools Help

Production Requests

- Non-Stream PRs
  - Clouds
    - CER4.1-4.2P4
      - CER4.1-4.2P4 Demo 01
    - CER4.1-4.3P3
      - CER4.1-4.3P3 Demo 01
  - Inversion
    - CER4.5-6.0P6 (PGE Disabled)
    - CER4.5-6.0P6 Terra-MODIS 14 (
    - CER4.5-6.2P3
      - CER4.5-6.2P3 Test PR FM3
    - CER4.5-6.4P2
      - CER4.5-6.4P2 Test PR FM2
  - SARB
    - CER7.2.1P2
      - CER7.2.1P2 Terra-Aqua-MODIS 1319-
      - CER7.2.1P2 Terra-MODIS 1318-15
  - TISA Averaging
    - CER7.3.1P2
      - CER7.3.1P2 Aqua-MODIS 1304-15
    - CER7.3.1P3
      - CER7.3.1P3 Terra-Aqua-MODIS 1308-
      - CER7.3.1P3 Terra-MODIS 1308-15
    - CER7.3.1P4
      - CER7.3.1P4 Terra-Aqua-MODIS 1313-
      - CER7.3.1P4 Terra-MODIS 1312-15
    - CER8.1P2
      - CER8.1P2 Terra-MODIS 1323-15

2005 -> 01 CER4.5-6.2P3 Test PR FM3 Last Refresh: 22:00:36

Instance	CATALYST Job Status	Science Status	Epilog Status	Deleted
20050101	completed	SUCCESS	SUCCESS	<input type="checkbox"/>
20050102	completed	SUCCESS	SUCCESS	<input type="checkbox"/>
20050103	completed	SUCCESS	SUCCESS	<input type="checkbox"/>
20050104	completed	SUCCESS	SUCCESS	<input type="checkbox"/>
20050105	failed_science	PGE_ERROR		<input type="checkbox"/>
20050106	failed_science	PGE_ERROR		<input type="checkbox"/>
20050107	failed_science	PGE_ERROR		<input type="checkbox"/>
20050108	failed_science	SIGSEGV		<input type="checkbox"/>
20050109	failed_science	PGE_ERROR		<input type="checkbox"/>
20050110	running			<input type="checkbox"/>
20050111	running			<input type="checkbox"/>
20050112	running			<input type="checkbox"/>
20050113	running			<input type="checkbox"/>
20050114	failed_epilog	SUCCESS	VERIFY_ERROR	<input type="checkbox"/>
20050115	running			<input type="checkbox"/>
20050116	waiting	SUCCESS	SUCCESS	<input type="checkbox"/>
20050117	waiting	SUCCESS	SUCCESS	<input type="checkbox"/>
20050118	waiting	SUCCESS	SUCCESS	<input type="checkbox"/>
20050119	waiting	SUCCESS	SUCCESS	<input type="checkbox"/>
20050120	epilog_running	SUCCESS		<input type="checkbox"/>
20050121	epilog_queued	SUCCESS		<input type="checkbox"/>
20050122	epilog_running	SUCCESS		<input type="checkbox"/>
20050123	waiting			<input type="checkbox"/>
20050124	waiting			<input type="checkbox"/>
20050125	waiting			<input type="checkbox"/>
20050126	waiting			<input type="checkbox"/>
20050127	queued			<input type="checkbox"/>
20050128	ready_to_launch			<input type="checkbox"/>
20050129	queued			<input type="checkbox"/>
20050130	waiting			<input type="checkbox"/>
20050131	waiting			<input type="checkbox"/>

This CATALYST job's science process has made an invalid memory reference.

16.13% FAILED

29.03% completed

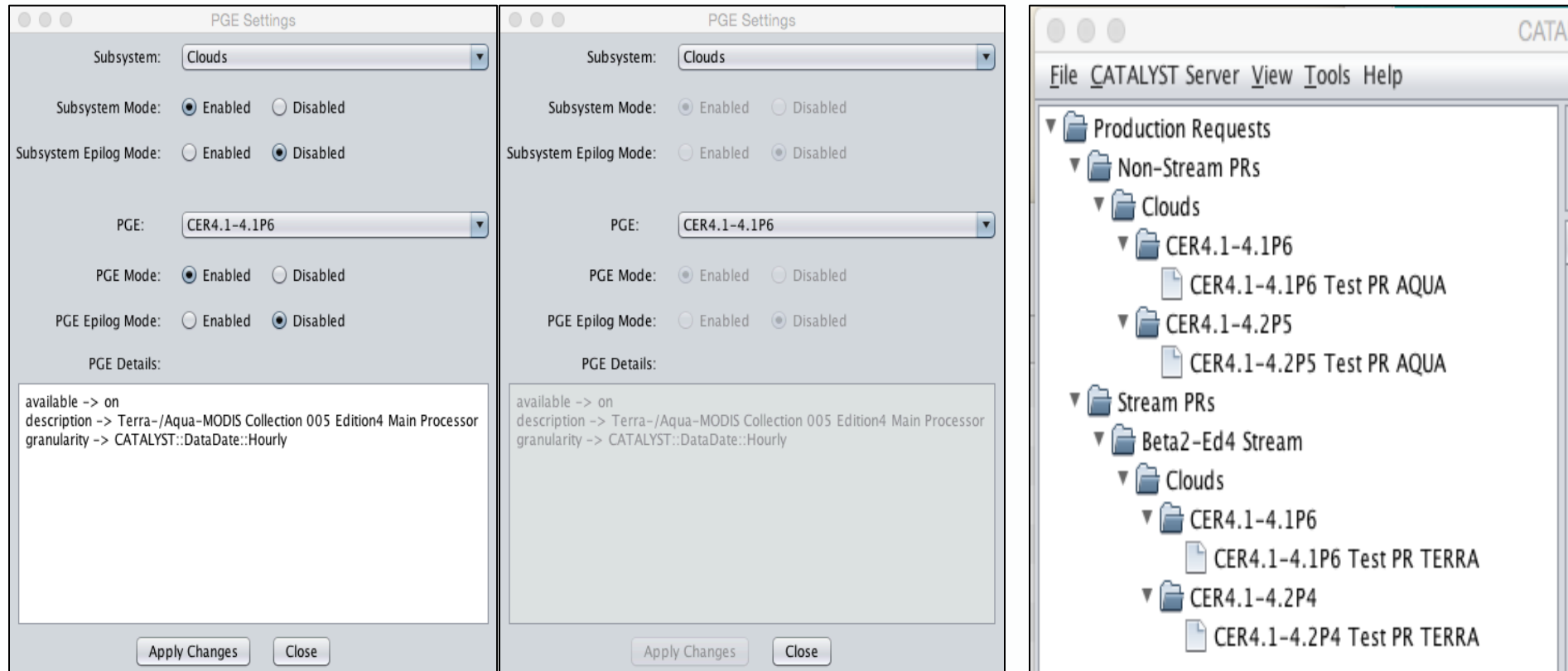
■ - Failed ■ - Success



# CATALYST 2.0 Improvements

## Operator's Console Improvements Cont'd

- PRs organized by processing stream
- Console window is resizable (previously static)
- Intelligent button disabling based on permissions



Functions grayed out if insufficient user permissions



# CATALYST 2.0 Improvements

## Stability & Monitoring

- Added auto-detection for inactive jobs in SGE grid engine
  - Detects if job manually deleted (“qdel-ed”) SGE jobs
  - Detects if blade issue causes job to hang
- New management utilities to monitor, start & stop CATALYST processes (Admin user only)
  - Graphical utility within Operator’s Console UI
  - Stand alone command line tool
- Improved system health monitoring email reporting
  - Generated by server process (previously standalone script)
  - More detailed and instructive info in reports – color coded to indicate status



# CATALYST 2.0 Improvements

## Logging

- Significantly improved logging capability
  - CATALYST Processes-level logging with all entries time stamped
  - Job details window now gives access within Console to:
    - PCF
    - PCF logs
    - LogStatus files
    - SGE logs
    - Epilogue logs

## Security

- SSL support enabled for CATALYST API
  - All comm. From PR Tool & Operator's Console now encrypted end-to-end
  - No visible changes to user